THE (MIS)REMEMBRANCE OF THINGS PAST: MECHANISMS OF MEMORY STORAGE, UPDATING, AND WHY WE MISREMEMBER

Memory is a critical function of the brain; we treasure many of our memories, and it is widely believed that our past experiences make us who we are. However, decades of psychological research has revealed that we are prone to having misinformation introduced into our memories, and a recent study has suggested that many people's 'first memories' are not actually real, but reconstructions based upon family stories and old photographs. So, how are memories stored in the brain, and how can it be that what we remember is not necessarily what actually happened?

Memory is a fascinating function of the brain, and one that has united psychologists, neuroscientists, physiologists and biochemists in trying to understand its underlying mechanisms. Memory can be studied at many levels, with the unifying view that memories are stored as 'traces' or 'engrams' within the brain following activation of specific networks of brain cells (neurons) during an experience. There are various types of memories, from memories of events that we can pass on in words, to learned emotional responses (e.g. fear) that we have to particular environmental cues (e.g. spiders). These memories upon different brain structures, and all depend upon the formation of engrams.

Over the past 50 years, our understanding of the mechanisms underlying memory has advanced markedly. It is recognised that biochemical changes occur at the level of individual neurons, increasing their signalling efficacy with other neurons within the memory trace. This view has traditionally emphasised the stability of memory. However, more recent work has revealed how we can balance stability and flexibility, allowing memories to be updated with new information. Memories, it seems, can be modified, allowing updating, but also inaccuracies to be introduced (whilst also providing an opportunity for developing new memory-based treatments for mental health disorders).

Memories are made of this: the synaptic plasticity and memory hypothesis

The search for the engram began in the 1900s with physiologists such as Richard Semon and Karl Lashley attempting to find memory traces in the brain. The notion of the engram entered mainstream memory research in 1949, when Donald Hebb proposed his theoretical mechanism of how learning could occur in the brain in his book, *The Organization of Behavior*. Now known as 'Hebb's rule', this encapsulates memory under the slogan "neurons that fire together, wire together". However, it was another 20 years before any evidence emerged to support Hebb's hypothesis.

Tim Bliss and Terje Lømo, working in the lab of Per Andersen at the University of Oslo, were investigating the effects of electrical stimulation of a specific brain structure, the hippocampus, in anaesthetised rabbits. What they discovered, and published in their seminal 1973 paper, was a process that they termed 'long-term potentiation' (LTP). This physiological process produced a long-lasting change in the efficiency of signalling between neurons in a network following a strong electrical stimulation. Further research revealed that LTP had properties that were consistent with associative learning: it required a level of stimulation beyond what is required for normal synaptic transmission (i.e. plasticity was 'cooperative'). LTP only worked when the two neurons involved were active within a brief time window (i.e. it was both associative, and input specific). These properties mapped well with the properties of associative learning, and led Bliss and Lømo to tentatively speculate that they may have found the mechanism by which Hebb's rule worked in the brain.

From 1973, the numbers of papers on LTP, particularly attempting to relate LTP to the process of memory consolidation, increased markedly. Decades of work by scientists prominent in the field – James McGaugh, Eric Kandel, Richard Morris, Graham Collingridge, to name but a few – has shown that both LTP and memory consolidation depend on a set of shared synaptic plasticity mechanisms (**Fig. 1**). These involve the NMDA subtype of postsynaptic ionotropic

glutamate receptor, the entry of calcium into the postsynaptic neuron, activation of protein kinases and ultimately the activation of gene transcription and protein synthesis to stabilise the structural changes that underlie the functional change in synaptic signalling efficacy.

The 'synaptic plasticity and memory' (SPM) hypothesis gained huge momentum in neuroscience, and focused on addressing whether changes in synaptic plasticity could be observed when an animal had undergone a learning experience (they can) and whether manipulating synaptic plasticity – blocking or overwriting it – could interfere with memory consolidation (it does). This was originally addressed through pharmacological studies, but more recent work from the labs of Susumu Tonegawa, Sheena Josselyn and Paul Frankland have begun to directly visualise and selectively manipulate the neurons that make up the memory trace. However, it is only recently that the strongest test of the SPM hypothesis has begun to be addressed experimentally. If our understanding of the molecular mechanisms of memory is correct, then it should be possible, according to the SPM view, to implant a memory by directly manipulating synaptic plasticity in the brain. Tim Bliss refers to this memory mimicry as his 'Marilyn Monroe' criterion, in light of the fact that he would like to have an implanted memory of having had dinner with the actress.

Experimentally, we are not *quite* there yet, but the recent advances in optogenetic technology have made this criterion more addressable. In mice, it is possible to implant a false memory of an aversive event occurring in a particular context. In 2013, the Tonegawa lab selectively labelled neurons involved in encoding the memory of one context (A) with a fluorescent marker and the light-sensitive channel Channelrhodopsin 2 (ChR2). They later exposed the mice to the aversive event of receiving a mild electric footshock in a distinct context (B), while simultaneously stimulating the labelled neurons with blue light. They found that when the mice were later tested, they showed fear to context A, even though they had never experienced an

aversive event in that context – essentially, they had artificially linked the memory of context A with the shock, and the mice behaved as if they really *had* experienced an aversive event in that environment. Therefore, even the strongest test of the SPM hypothesis is experimentally supported. Memory appears to be based on long-term changes in synaptic plasticity occurring between networks of neurons that form the engram.

Keeping it real(?): the reconsolidation hypothesis

The emphasis of memory research in neuroscience, however, presents a problem; if memories are based on long-term, stable changes in synaptic plasticity, then how can we account for the psychological findings suggesting that memories are often inaccurate, and prone to the introduction of misinformation? Psychologists have traditionally placed emphasis on the *dynamic* nature of memory, rather than the stability of the engram emphasised by neuroscientists. How to resolve these two different points of view?

Even as far back as the 1960s, there had been indications from the neuroscience literature that the view emphasising memory stability may not be completely accurate. In 1968, two articles published in *Science* suggested that even old, well-established memories become malleable under certain conditions of retrieval. However, failure of prominent memory researchers to replicate these findings, and the lack of a molecular mechanism to account for the behavioural findings, made 'cue-dependent amnesia' relatively easy to dismiss. For thirty years, the field instead focused, with great productivity and success, on characterising the molecular mechanisms of LTP and memory consolidation.

However, in 2000, evidence emerged that made the 'cue-dependent amnesia' findings harder to ignore. Karim Nader, working in the lab of Joseph LeDoux, showed that a fully consolidated fear memory could, under the right conditions of retrieval, return to a state where it was sensitive to manipulation with an amnestic agent. Nader's account was clearer than the original papers. He tested the memory of pavlovian fear conditioning, which was well-characterised psychologically and

in terms of the underlying brain circuitry. Nader targeted a brain structure, the amygdala, known to be critical for the storage of the pavlovian fear memory trace, and induced amnesia using a protein synthesis inhibitor, which had a recognised mechanism of action and could be related to synaptic plasticity changes. The 'cue-dependent amnesia' research was therefore rediscovered, reinvigorated, and renamed as 'memory reconsolidation'.

The reconsolidation view has similarities to the original 'consolidation' view of memory, but with some important differences (**Fig. 2**) that allow memory to be seen as much more dynamic and flexible than the traditional view. The key difference is that, rather the considering a memory as existing in a short-term or long-term store, it should be considered in terms of its activity state. Memories in the active state are transient and highly malleable – much like the traditional view of short-term memory, whilst memories in the inactive state are more stable and persistent, similar to the traditional view of long-term memory. The critical difference in the reconsolidation view is that memories do not simply move from a short-term to long-term memory store, but instead can cycle between active and inactive states throughout their lifespans. Under certain conditions of retrieval – and it is not necessarily the case that every retrieval event converts an inactive memory to the malleable active state – memories are destabilised and returned to a state in which they can become updated. This capacity for updating also makes the memory vulnerable; similar to when you are drafting an email and the computer crashes before it is sent. The ability to edit and redraft is highly adaptive, but under specific (unfortunate) conditions it would be possible to lose what was written.

Reconsolidation: not just a rerun of consolidation

Although initially met with some scepticism, over the past two decades the reconsolidation hypothesis has become increasingly accepted by the field. Reconsolidation has been observed by multiple labs, investigating different types of memory, in species ranging from crabs to humans. The research has tended to focus in two different, though related, directions: one basic line of research has been attempting to characterise the molecular mechanisms of reconsolidation, and comparing

them to consolidation, while the other, more applied, line of research has been attempting to determine whether reconsolidation could be exploited to treat mental health disorders such as post-traumatic stress disorder and drug addiction.

The processes by which a memory reconsolidates (or, more mechanistically, restabilises) are similar to those that underlie initial memory consolidation (**Fig. 1**). Reconsolidation depends upon activity at NMDA receptors, the activation of protein kinases, and the recruitment of transcription factors to initiate protein synthesis in the same way as consolidation, though there do appear to be differences in the pathways underlying the two processes. (For example, while the consolidation but not reconsolidation of a contextual fear memory depends upon the protein BDNF, reconsolidation but not consolidation of the same memory depends upon the protein Zif268.) The processes differ more with respect to the mechanisms underlying the conversion of the memory from the inactive to the active state; destabilisation, which is specific to reconsolidation (**Fig. 2**). This appears to require protein degradation through the proteasome system, possibly depending upon similar mechanisms to processes underlying synaptic weakening. At the cell-surface level, destabilisation appears to be activated by particular receptors, many of which modulate intracellular signalling either through specific patterns of calcium activation (L-type voltage-gated calcium channels and NMDA receptors containing the GluN2B subunit) or by activating intracellular second messengers.

One question of intense interest has been the requirement for the neurotransmitter dopamine; release of this neurotransmitter has previously been shown to correlate with situations in which there is a violation of expectations, or 'prediction error'. Intuitively, it would make sense that memories would enter a state in which they could be updated in situations where an individual experiences surprise; using the memory produces a prediction that is incorrect. This would indicate that the memory was not fully accurate, and needs to be updated. This view has been supported with behavioural evidence from crabs, rats and humans – memories update when there is new information to be incorporated – and research is ongoing to determine how dopaminergic signalling relates to this.

Exploiting reconsolidation to treat mental health disorders

As noted above, a major implication of the reconsolidation view is that even old, fully consolidated memories can return to an unstable and malleable state under the right conditions. This has a potential impact in the development of new therapies for mental health disorders such as post-traumatic stress disorder (PTSD), phobia, and even drug addiction. In these disorders, emotional memories associating environmental cues (e.g. spiders, syringes) with motivationally relevant outcomes (e.g. fear, a drug high) become maladaptive and come to dominate behaviour. For example, a patient with severe spider phobia may refuse to leave the house in case they encounter a spider. These memories are usually old and well-established before a patient presents in the clinic looking for treatment, so the possibility of inducing the maladaptive memory into an unstable state in which it can be disrupted pharmacologically (as has been done by the lab of Merel Kindt) or interfered with by the use of specific cognitive tasks (such as the computer game Tetris, in the lab of Emily Holmes) is very promising for new treatment development. The more understanding of the basic mechanisms develops, the more refined these treatments will become.

Final thoughts on memory: looking forward, not back

Though the capacity to update gives us a dynamic and flexible memory system, the downside is that this allows for misinformation to be introduced into memories. Studies in psychology, including the seminal work of Elizabeth Loftus, has made it clear that memories are not 100% accurate, and that information presented after an event – for example, through the use of misleading questions – can affect the way in which an event is remembered. This may seem like a major evolutionary disadvantage, but in reality it is a relatively small price to pay for a memory system that can update with more recent and relevant information. As Daniel Schacter noted when considering his '7 sins of memory', it is important to appreciate that the function of memory is not to look back, but forward; in evolutionary terms, our memories need only to be accurate enough to know how we should

behave the next time we are in a similar situation, not to recall with perfect accuracy what happened in the past.

Further reading

- Akhtar S, Justice LV, Morrison CM & Conway MA (2018) Fictional first memories. Psychological Science (*in press).* doi: 10.1177/0956797618778831.
- Fenno L, Yizhar O & Deisseroth K (2011) The development and application of optogenetics. Annual Review of Neuroscience 34: 389-412.
- Josselyn SA, Köhler S & Frankland PW (2015) Finding the engram. Nature Reviews Neuroscience 16: 521-534.
- Loftus EF (2017) Eavesdropping on memory. Annual Review of Psychology 68: 1-18.
- Martin SJ & Morris RGM (2002) New life in an old idea: the synaptic plasticity and memory hypothesis revisited. Hippocampus 12: 609-636.
- Milton AL & Everitt BJ (2010) The psychological and neurochemical mechanisms of drug memory reconsolidation: implications for the treatment of addiction. European Journal of Neuroscience 31: 2308-2319.
- Nicoll RA (2017) A brief history of long-term potentiation. Neuron 93(2): 281-290.
- Ramirez S, Liu X, Suh J, Pignatelli M, Redondo RL, Ryan TJ & Tonegawa S (2013) Creating a false memory in the hippocampus. Science 341(6144): 387-391.
- Schacter DL (1999) The seven sins of memory: insights from psychology and cognitive neuroscience. American Psychology 54: 182-203.
- Tronson NC & Taylor JR (2007) Molecular mechanisms of memory reconsolidation. Nature Reviews Neuroscience 8: 262-275.
- Visser RM, Lau-Zhu A, Henson RN & Holmes EA (2018) Multiple memory systems, multiple time points: how science can inform treatment to control the expression of unwanted

emotional memories. Philosophical Transactions of the Royal Society B 373(1742). pii: 20170027.

Figure legends

Fig. 1. Simplified schematic of the post-synaptic molecular mechanisms underlying long-term potentiation and memory consolidation. The NMDA subtype of glutamate receptor is constitutively blocked by a magnesium ion, which is released when the post-synaptic neuron is sufficiently depolarised (e.g. by activation of the AMPA subtype of glutamate receptor, which is permeable to Na^+). Thus, the NMDA receptor is double-gated, requiring both post-synaptic depolarisation and glutamate binding to allow Ca^{2+} into the cell. This calcium influx initiates a signalling cascade that ultimately results in the synthesis of new proteins to stabilise structural changes that support the functional changes in signalling efficiency. Abbeviations: AC, adenylyl cyclase; CaM, calmodulin; CaMK, calcium-calmodulin-dependent protein kinase; cAMP, cyclic adenosine monophosphate; CRE, cAMP response element; CREB, cAMP response element binding protein; ERK, extracellular signal-regulated kinase; GPCRs, G-protein coupled receptors; MEK, mitogen-activated protein kinase kinase; PKA, protein kinase A.

Fig. 2. Schematic of the reconsolidation hypothesis. Rather than defining memories as being in 'short-term' or 'long-term' memory stores, it is more accurate to characterise them as being in malleable, unstable 'active' state, or a more stable and persistent 'inactive' state. According to the reconsolidation hypothesis, memories can move from the inactive to the active state under certain conditions of retrieval (likely involving a violation of expectations) and consequently reconsolidate (or, more mechanistically, restabilise) back to the inactive state in the updated form.

Biographical information

Amy Milton is a Lecturer in Experimental Psychology at the University of Cambridge, and the Ferreras-Willetts Fellow in Neuroscience at Downing College, Cambridge. Her research interests focus on the neurochemical and molecular mechanisms of memory reconsolidation, and the application of reconsolidation-based interventions to mental health disorders including post-traumatic stress disorder, drug addiction, and obsessive-compulsive disorder. Email: alm46@cam.ac.uk

Fig. 1

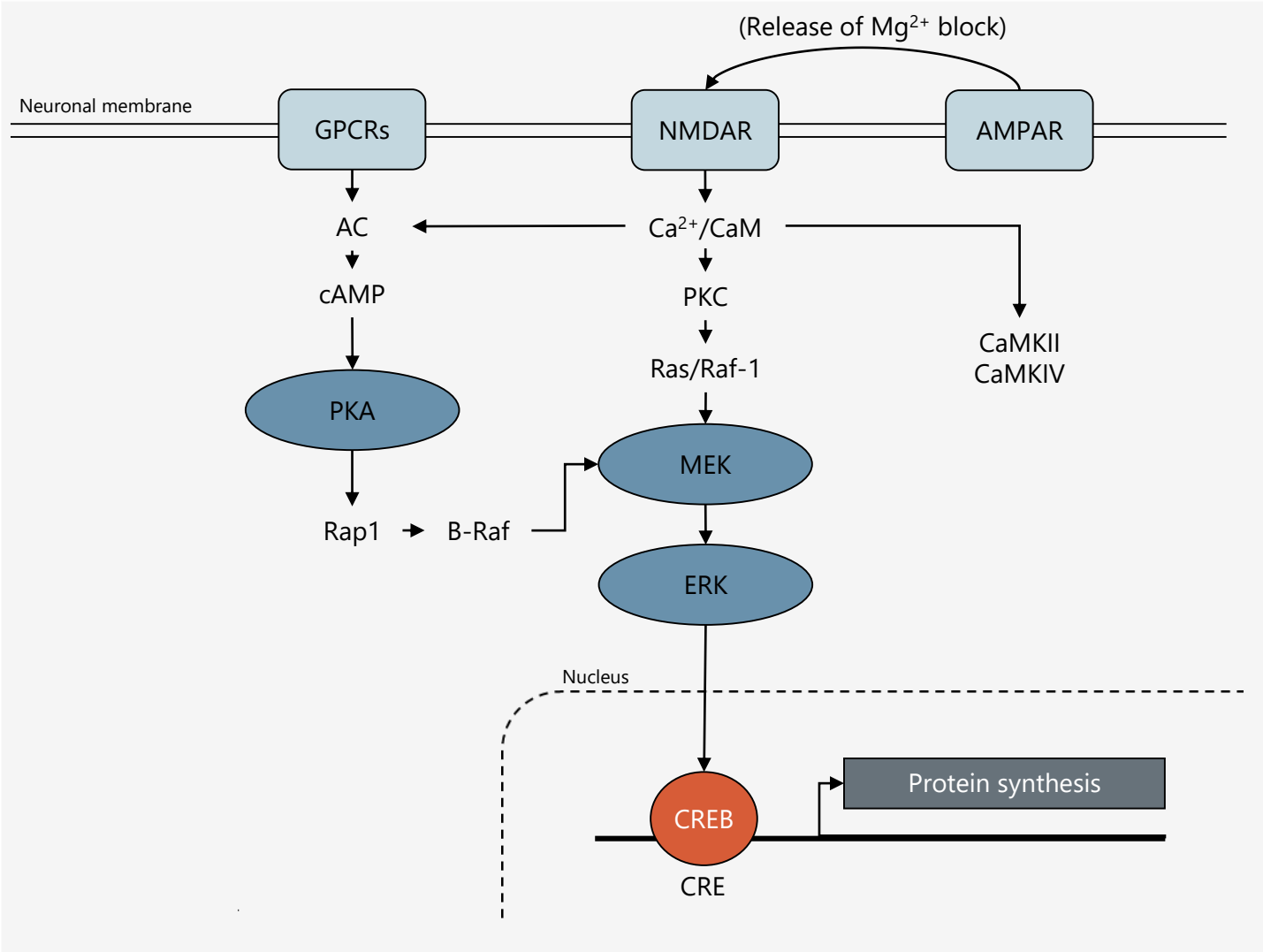

Fig. 1. Simplified schematic of the post-synaptic molecular mechanisms underlying long-term potentiation and memory consolidation. The NMDA subtype of glutamate receptor is constitutively blocked by a magnesium ion, which is released when the post-synaptic neuron is sufficiently depolarised (e.g. by activation of the AMPA subtype of glutamate receptor, which is permeable to Na^+). Thus, the NMDA receptor is double-gated, requiring both post-synaptic depolarisation and glutamate binding to allow Ca^{2+} into the cell. This calcium influx initiates a signalling cascade that ultimately results in the synthesis of new proteins to stabilise structural changes that support the functional changes in signalling efficiency. Abbeviations: AC, adenylyl cyclase; CaM, calmodulin; CaMK, calcium-calmodulin-dependent protein kinase; cAMP, cyclic adenosine monophosphate; CRE, cAMP response element; CREB, cAMP response element binding protein; ERK, extracellular signal-regulated kinase; GPCRs, G-protein coupled receptors; MEK, mitogen-activated protein kinase kinase; PKA, protein kinase A.

Fig. 2

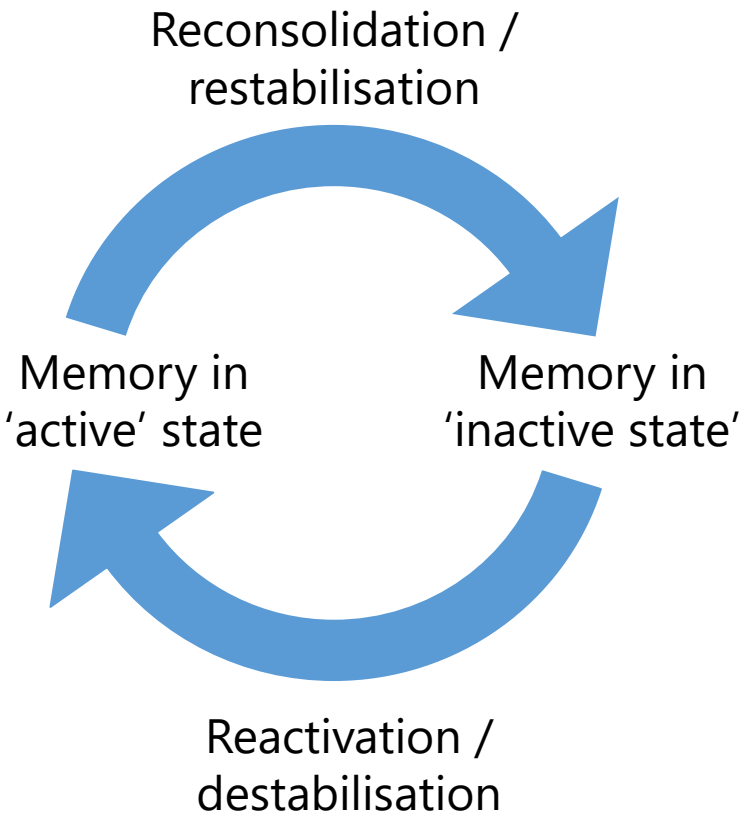

Fig. 2. Schematic of the reconsolidation hypothesis. Rather than defining memories as being in 'short-term' or 'long-term' memory stores, it is more accurate to characterise them as being in malleable, unstable 'active' state, or a more stable and persistent 'inactive' state. According to the reconsolidation hypothesis, memories can move from the inactive to the active state under certain conditions of retrieval (likely involving a violation of expectations) and consequently reconsolidate (or, more mechanistically, restabilise) back to the inactive state in the updated form.